UP TO NOW...

- Chapter 2: Types of Data (Nominal/Ordinal/Interval/Ratio, Discrete/Continuous), Descriptive Statistics (Central Tendency, Variation), Graphical Display (Histogram, Stem-and-Leaf, Box Plot)

- Chapter 3: Probability (Classical, Relative Frequency, Subjective), Type of Events (Mutually Exclusive, Independent), Laws (Addition, Multiplication), Conditional Probability, Medical Tests (Sensitivity, Specificity, Predictive Value)

PREVIOUSLY ...

Use sample data to compute the probability of an event

FROM NOW ON ...

the above plus …

Make decisions using sample data based on difference(s) between computed from expected probabilities given an assumed probability distribution

DEFINITIONS

- ***Random Variable***

A numeric *function* that assigns probabilities to different events in a sample space. (Rosner)

A numerically valued *quantity* that takes on specific values with different probabilities. (Rosner - Study Guide)

A *variable* whose values occur by chance and cannot be predicted exactly in advance. (Daniel)

A *variable* that has a single numerical value, determined by chance, for each outcome of a procedure. (Triola)

The numerical *outcome* of a random experiment. (Gonick - Cartoon Guide)

A variable whose possible values are numerical outcomes of a random phenomenon. (Yale Stat Web Site)

Discrete Random Variable

one with a finite set of values, or a countable set of values (Rosner refers to a discrete set of values with specified probabilities)

discrete random variables are usually counts

number of low birth weight infants, number of cases of breast cancer, number of heads when tossing a coin, sum when rolling two dice

Continuous Random Variable

one that has infinitely many values, and those values are associated with measurements on a continuous scale without gaps or interruptions (Rosner distinguishes continuous from discrete random variables by saying that the possible values of a continuous random variable cannot be enumerated)

continuous random variables are usually measurements

birth weight, height, weight, blood pressure

- ***Probability Distribution***

A *mathematical relationship* (or rule) that assigns to any possible value r of a discrete random variable X the probability P(X=r).

The *relationship* between the values and their associated probabilities. (Rosner - Study Guide)

> Note: Rosner calls both of the above probability-mass functions and then says that a probability-mass function is also referred to as a probability distribution.

A *graph, table, or formula* that gives the probability for each value of a random variable. (Triola)

A *device* that summarizes the relationship between the values of random variables and the probabilities of their occurrence. It may be expressed in the form of a *graph, table, or formula*. (Daniel)

Discrete Probability Distribution

Binomial, Poisson
Rosner Chapter 4

Continuous Probability Distribution

Normal
Rosner Chapter 5

for more on definitions, see Statistics Glossary

CLASSICAL PROBABILITY

some discrete events where we can create a create theoretical models (probability distributions)...

toss one coin many times
random variable X is the number of heads
probability distribution (T:0, H:1)

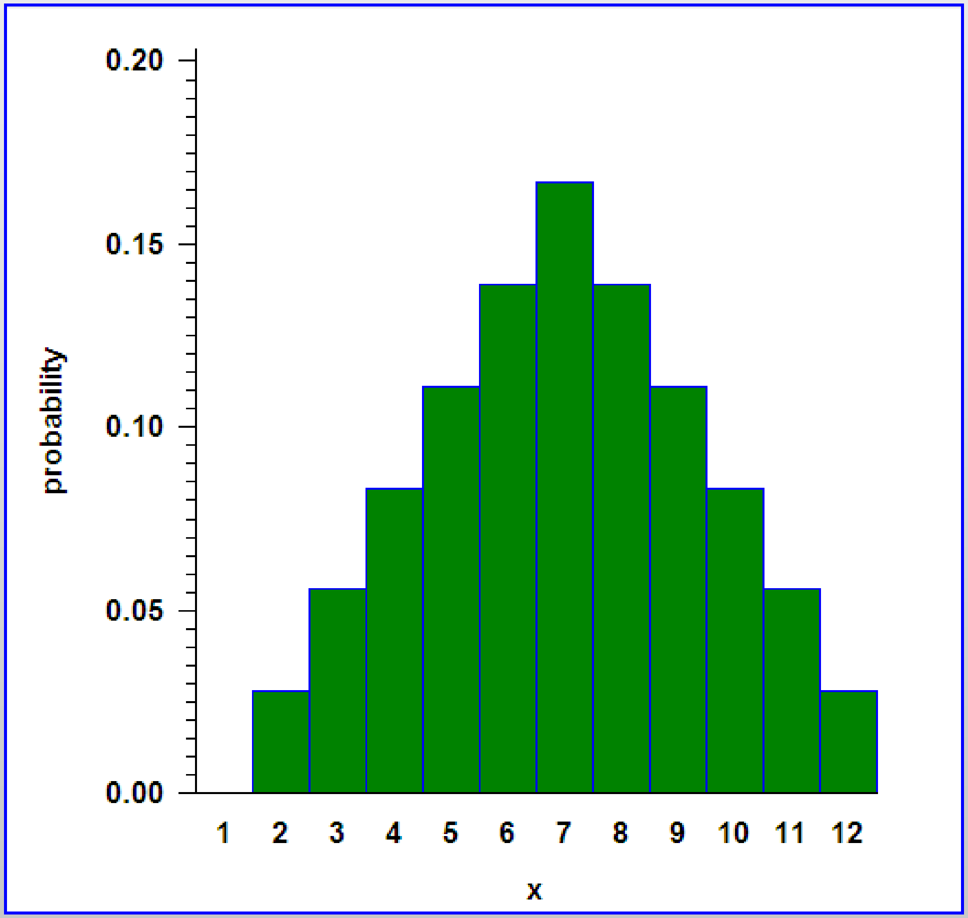
outcome(x)	0	1
P(X=x)	0.5	0.5

toss two coins many times
random variable X is the number of heads
probability distribution (TT:0, TH,HT:1, HH:2)

outcome(x)	0	1	2
P(X=x)	0.25	0.5	0.25

rolling two dice many times ...
discrete random variable X is the sum of the two dice
probability distribution is shown below

Probability Distribution	
x	P(X=X)
2	1/36 = 0.028
3	2/36 = 0.056
4	3/36 = 0.083
5	4/36 = 0.111
6	5/36 = 0.139
7	6/36 = 0.167
8	5/36 = 0.139
9	4/36 = 0.111
10	3/36 = 0.083
11	2/36 = 0.056
12	1/36 = 0.028



RELATIVE FREQUENCY PROBABILITY

example 4.6 from Rosner ...

4 patients take a drug to control hypertension
random variable X is the number of patients with controlled hypertension (probabilities said to be from experience)

outcome(x)	0	1	2	3	4
P(X=x)	0.008	0.076	0.265	0.411	0.240

note all probabilities are in the range 0 to 1 and that probabilities sum to 1 (true for any probability-mass function)

example 4.8 from Rosner …

100 MDs use the drug on 4 patients and record outcomes

expected …

outcome(x)	0	1	2	3	4
P(X=x)	0.008	0.076	0.265	0.411	0.240

study results … *proportion* of MDs reporting number of patients with controlled hypertension …

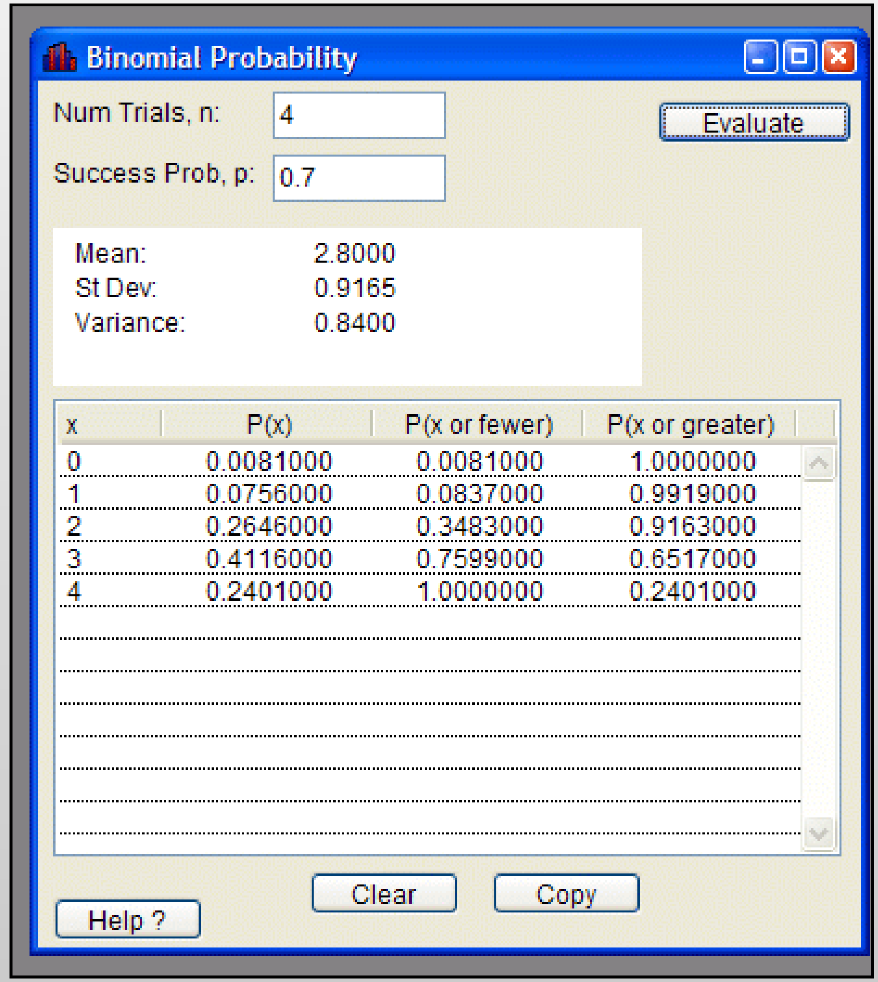
outcome(x)	0	1	2	3	4
P(X=x)	0.000	0.090	0.240	0.480	0.190

are study results different from expected results what is the source of the expected results

sometimes the source is already existing data

sometimes the source is a probability distribution

in example 4.8, the source was the binomial distribution with 4 trials and a probability of success of 0.7 (column P(x) on the right)



EXPECTED VALUE OF A DISCRETE RANDOM VARIABLE

$E(X) = \mu = \sum x\, P(X=x)$

x's are values that the random variable assumes

E(X) for rolling two dice
P(X=x) from table on page 9 ...

$$E(X) = (2 \times 0.028) + (3 \times 0.056) + (4 \times 0.083) + (5 \times 0.111) + (6 \times 0.139) + (7 \times 0.167) + (8 \times 0.139) + (9 \times 0.111) + (10 \times 0.083) + (11 \times 0.056) + (12 \times 0.028)$$

$E(X) = \mu = 7$

note: just a sum of possible outcomes weighted by the probability of each outcome

example 4.8 from Rosner …

E(X) for controlled hypertension
P(X=x) from table on page 10 …

$$E(X) = (0 \times 0.008) + (1 \times 0.076) + (2 \times 0.265) + (3 \times 0.411) + (4 \times 0.240)$$

$E(X) = \mu = 2.80$ (look at mean in table in page 12)

compute mean from sample data …

$$\overline{X} = (0 \times 0.000) + (1 \times 0.090) + (2 \times 0.240) + (3 \times 0.480) + (4 \times 0.190)$$

$$\overline{X} = 2.77$$

are study results different from expected results

VARIANCE OF A DISCRETE RANDOM VARIABLE

$\text{Var}(X) = \sigma^2 = \sum (x - \mu)^2 \Pr(X=x)$

x's are values that the random variable assumes

Var(X) for rolling two dice...

$$\begin{aligned}\text{Var}(X) = {} & (2-7)^2 \times 0.028 + (3-7)^2 \times 0.056 + (4-7)^2 \times 0.083 + \\ & (5-7)^2 \times 0.111 + (6-7)^2 \times 0.139 + (7-7)^2 \times 0.167 + \\ & (8-7)^2 \times 0.139 + (9-7)^2 \times 0.111 + (10-7)^2 \times 0.083 + \\ & (11-7)^2 \times 0.056 + (12-7)^2 \times 0.028\end{aligned}$$

$\text{Var}(X) = \sigma^2 = 5.83$ Standard Deviation = 2.42

- what is a rare event given a roll of two dice
- what is the probability of getting a 12
- if you think of 95% of values being between the mean ± 2 standard deviations, what is that range of values
- is a 12 rare? is a 2 rare? is a 3 rare?

---

rare events often thought of as those with a probability < 0.05

$P(X=12) = 0.028$

$\mu \pm 2\sigma = 7 \pm 2 \times 2.42$ 95% of values in range 2.16 - 11.8

example from Triola...

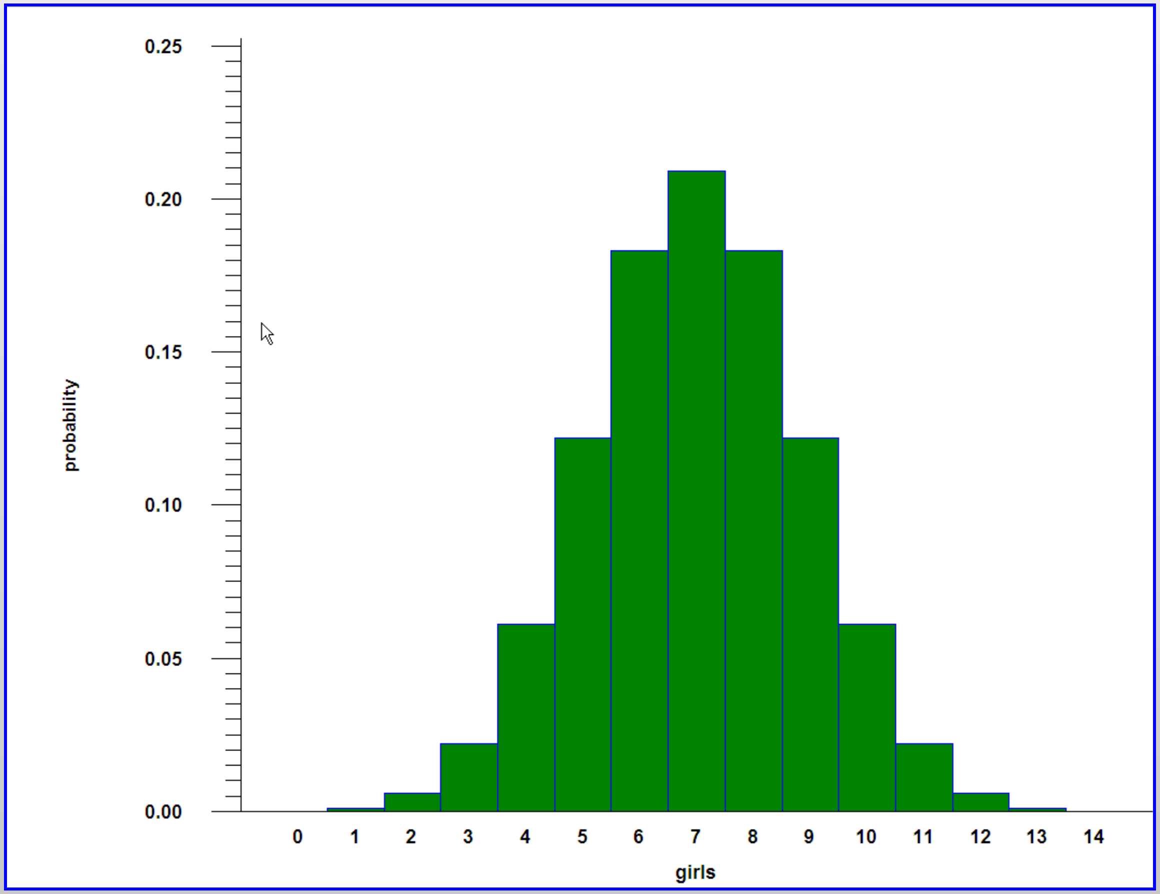
IVF method developed that increases the probability of a couple having a girl … among 14 couples, 13 have a girl

is this enough evidence to conclude that the new method is effective

or …

is 13 girls out of 14 births a rare event need the probability distribution ...

Probability Distribution	
girls	P(X=girls)
0	0.000
1	0.001
2	0.006
3	0.022
4	0.061
5	0.122
6	0.183
7	0.209
8	0.183
9	0.122
10	0.061
11	0.022
12	0.006
13	0.001
14	0.000



$E(X) = \mu = \sum x \, Pr(X=x)$

$Var(X) = \sigma^2 = \sum (x - \mu)^2 \, Pr(X=x)$

$\mu = 6.993 \approx 7.0$

$\sigma^2 = 3.564 \approx 3.6$

$\sigma = 1.9$

$\mu \pm 2\sigma = 7.0 \pm 2 \times 1.9$ 95% of values in range 3.2 - 10.8

$P(X \geq 13) < .05$

13 is also outside the mean ± 2 standard deviations

conclusion?

given the table on page 18, how many girls would make you think that the IVF method results in more girls

event	girls
GGGG	4
GGGB	3
GGBG	3
GBGG	3
BGGG	3
GGBB	2
BGGB	2
BBGG	2
GBBG	2
GBGB	2
BGBG	2
BBBG	1
BBGB	1
BGBB	1
GBBB	1
BBBB	0

WHERE DID THE PROBABILITIES COME FROM

reduce the problem to 4 couples
what is the sample space
number of events: $2^4 = 16$

each event has the same probability
P = 0.50 x 0.50 x 0.50 x 0.50 = 0.0625

using frequencies from the sample space on the left ...

Probability Distribution	
girls	P(X=girls)
4	0.0625
3	0.2500
2	0.3750
1	0.2500
0	0.0625

ANOTHER PROBLEM

The incidence of Down Syndrome increase with mother's age…

AGE	INCIDENCE
<30	<1 IN 1,000
30-34	1 IN 900
35	1 IN 400
36	1 IN 300
37	1 IN 230
38	1 IN 180
39	1 IN 135
40-41	1 IN 105
42-43	1 IN 60
44-45	1 IN 35
46-47	1 IN 20
48	1 IN 16
49	1 IN 12

According to the National Association for Down syndrome, "80% of babies born with Down syndrome are born to women younger than 35. The average maternal age is 28 years old." The likelihood of a woman under 30 years of age giving birth to a child with Down syndrome is less than 1:1000, but increases the older the woman gets, with an incidence of about 1:60 at 42 years of age.

event	ds
DDDD	4
DDDN	3
DDND	3
DNDD	3
NDDD	3
DDNN	2
NDDN	2
NNDD	2
DNND	2
DNDN	2
NDND	2
NNND	1
NNDN	1
NDNN	1
DNNN	1
NNNN	0

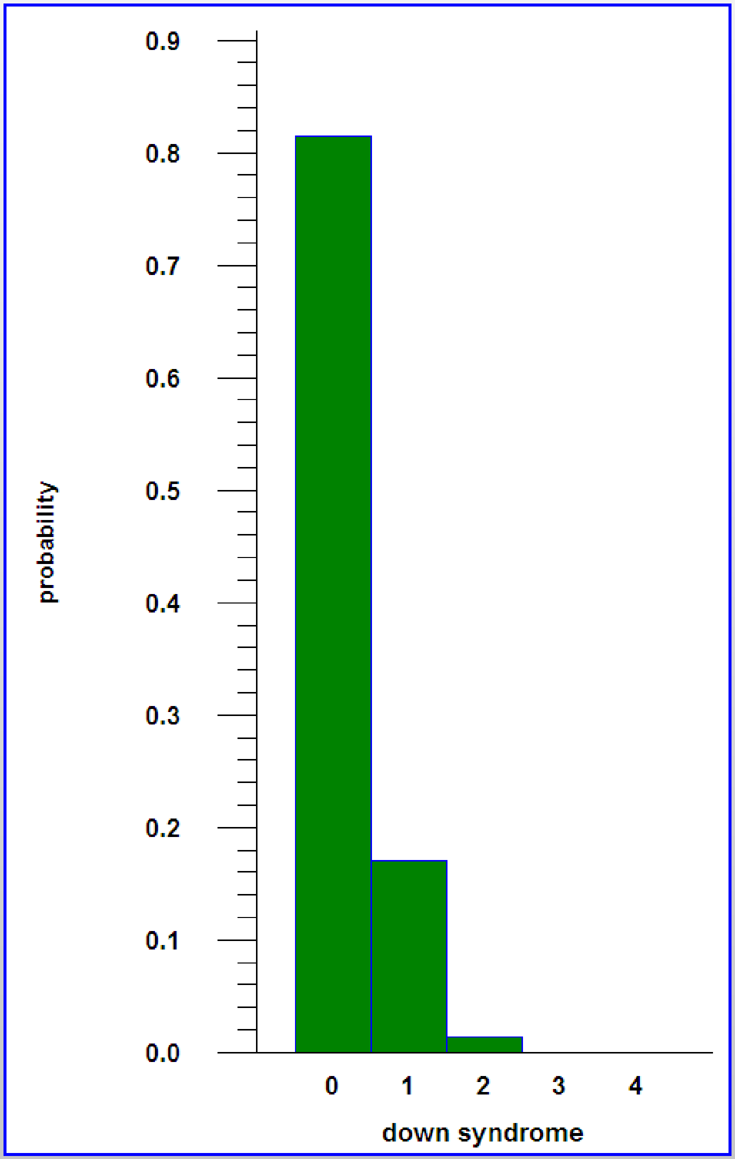
given 4 births among women age 46, what is the probability distribution of births with Down Syndrome (ds)
number of events ... $2^4 = 16$
from table ... at age 46 ... P(D) = 1/20 = 0.05

not like the IVF boy/girl problem since each event does not have the same probability ...
4 cases = 0.05 x 0.05 x 0.05 x 0.05 = 0.000
3 cases = 0.05 x 0.05 x 0.05 x 0.95 = 0.000
2 cases = 0.05 x 0.05 x 0.95 x 0.95 = 0.002
1 case = 0.05 x 0.95 x 0.95 x 0.95 = 0.043
0 cases = 0.95 x 0.95 x 0.95 x 0.95 = 0.815

Probability Distribution	
ds	P(X=ds)
4	0.000
3	0.000
2	0.014
1	0.171
0	0.815

when P≠0.50, the probability distribution is not symmetric ...

Probability Distribution	
ds	P(X=ds)
4	0.000
3	0.000
2	0.014
1	0.171
0	0.815



COMBINATIONS

- when the number of events is small, it is not too difficult to list all the possible events in the sample space

- the examples with number of girls in four births or the number of Down syndrome infants in four births had 16 possible events in the sample space

- with a binary outcome (girl / boy, Down syndrome yes / no), the number of events is 2^N, where N is the number of trials within each event

- after listing the sample space, it is not too difficult to determine the number of 'successes' (for example, girls) within each event and the number of events with a given number of successes given four births … how many combinations produce …

4 girls? 3 girls? 2 girls? 1 girl? 0 girls?

combination	# of girls	# of combinations
GGGG	4	1
GGGB, GGBG, GBGG, BGGG	3	4
GGBB, BGGB, BBGG, GBBG, GBGB, BGBG	2	6
BBBG, BBGB, BGBB, GBBB	1	4
BBBB	0	1

another way to look at the sample space with four births …

```
1ST BIRTH IS A GIRL                          1ST BIRTH IS A BOY

                          GGGG                                         BBBB
               GGG  ->     +                              BBB  ->       +
                |         GGGB                             |           BBBG
       GG  ->   +                                 BB  ->   +
        |       |         GGBG                     |       |           BBGB
        |      GGB  ->     +                       |      BBG  ->       +
        |                 GGBB                     |                   BBGG
G  ->   +                                 B  ->    +
        |                 GBGG                     |                   BGBB
        |      GBG  ->     +                       |      BGB  ->       +
        |       |         GBGB                     |       |           BGBG
       GB  ->   +                                 BG  ->   +
                |         GBBG                             |           BGGB
               GBB  ->     +                              BGG  ->       +
                          GBBB                                         BGGG
```

- when the number of trials is large (IVP example and determing the probability of 13 girls in 14 births), the number of events in the sample space is large (2^{14} = 16,384)

- when the number of trials is large, determining the number of events with a given number of successes is not easy to enumerate (how many events have exactly 8 girls)

- use the formula for COMBINATIONS --- how many ways can k items be selected from n items when different orderings of the same items are not counted separately

  NOTE: when different orderings of the same items are counted separately, the formula for PERMUTATIONS is used --- do not bother with PERMUTATIONS for now, they are not used in understanding any of the discrete or continuous probability distributions

COMBINATIONS

$$nCk = \frac{n!}{k!(n-k)!}$$

where... n is the number of trials
k is the number of successes

factorial notation... n! = n x (n-1) x ... x 2 x 1

given 5 births ... how many ways are there to have exactly 2 girls ...

$$_5C_2 = \frac{5!}{2!(5-2)!} = \frac{5\times 4\times 3\times 2\times 1}{2\times 1(3\times 2\times 1)} = \frac{120}{12} = 10$$

given 5 births and that the probability of low birth weight is 0.10, show the probability distribution of having 0 through 5 low birth weight infants ...

how many ways can you select k items (0 through 5 of low birth weight infants) from n items (in this case, 5 births) ...

k	5 births	same as	event probability	#events $_nC_k$	total probability
0	.9x.9x.9x.9x.9	$.1^0 \times .9^5$	0.59049	1	0.59049
1	.1x.9x.9x.9x.9	$.1^1 \times .9^4$	0.06561	5	0.32805
2	.1x.1x.9x.9x.9	$.1^2 \times .9^3$	0.00729	10	0.07290
3	.1x.1x.1x.9x.9	$.1^3 \times .9^2$	0.00081	10	0.00810
4	.1x.1x.1x.1x.9	$.1^4 \times .9^1$	0.00009	5	0.00045
5	.1x.1x.1x.1x.1	$.1^5 \times .9^0$	0.00001	1	0.00001

event probability assumes independent events, uses the multiplication rule
the sum of the #events column ($_nC_k$) is 32 (or 2^5)
the sum of the total probability column is 1

- the table is similar to work already done in previous problems
- rather that enumerate all possible events in the sample space, the combination formula is used to determine the number of events with a given number of success (k)
- the probability of a given event is multiplied by the number of times that given event occurs in the sample space.
- each entry in the total probability column can be seen to be...

$$P(X=k) = ({}_nC_k)\, p^k q^{n-k}$$

where... p = the probability of event k (success)
q = 1 - p

BINOMIAL PROBABILITY DISTRIBUTION

$$P(X = k) = ({}_nC_k)\, p^k q^{n-k}$$

random variable X is the number of occurrences of an event over some fixed number of trials ... k is the observed number of events in n trials ... p is the probability of an event (q is the complement of p)

... the binomial probability formula is applicable to situations that meet the following criteria ...

- fixed number of trials
- trials are independent
- each trial has only two possible outcomes
- probabilities remain constant for each trials

AN EXAMPLE OF USING THE FORMULA

given that the probability of low birth weight is 0.1 (p), if 10 (n) births occur, what is the probability that exactly 3 (k) of them will be low birth weight

is this a binomial problem ...

fixed number of trials	yes, n=10
trials are independent	yes, birth weight of one infant has no effect on the birth weight of another
each trial has only two possible outcomes	yes, low or normal birth weight
probabilities remain constant for each trials	yes, it is always 0.1

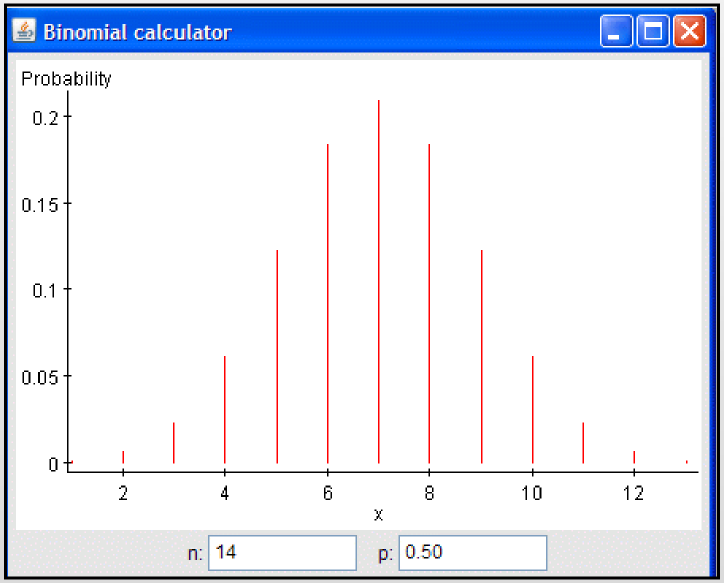
$$P(X = 3) = ({}_{10}C_3)0.1^3 0.9^7 = 120 \times 0.001 \times 0.478 = 0.057$$

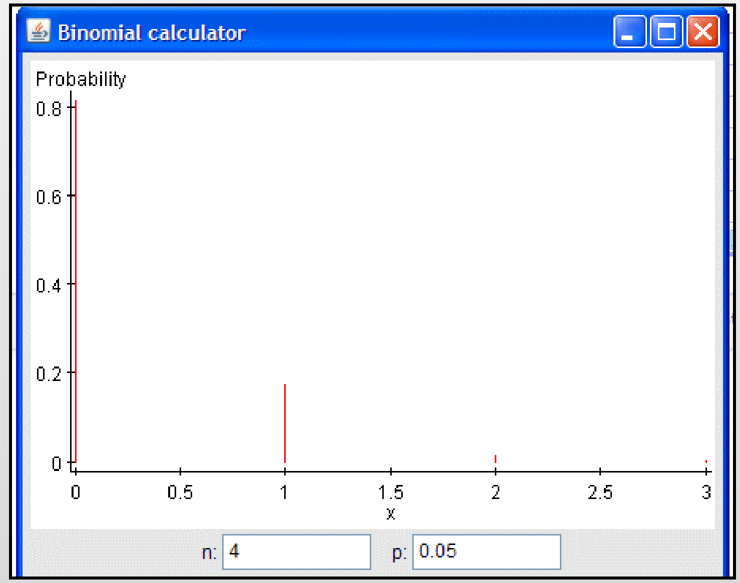
SOME BINOMIAL PROBABILITY DISTRIBUTIONS

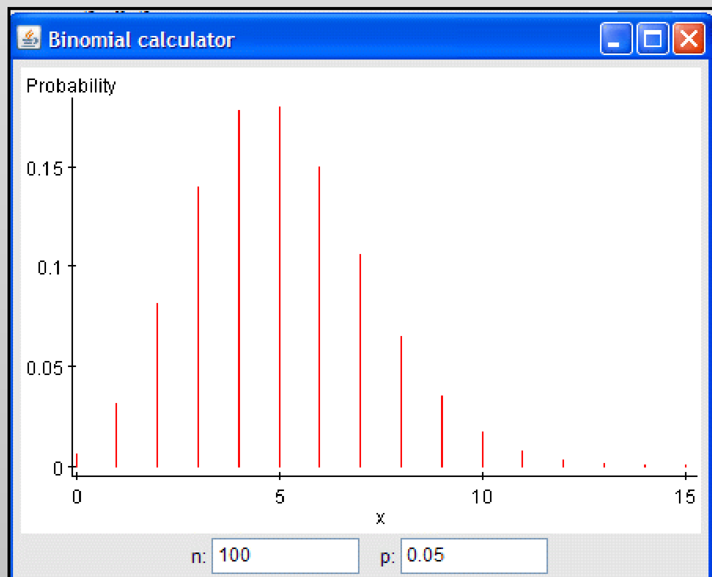
from previous examples ... always symmetric when p=0.5

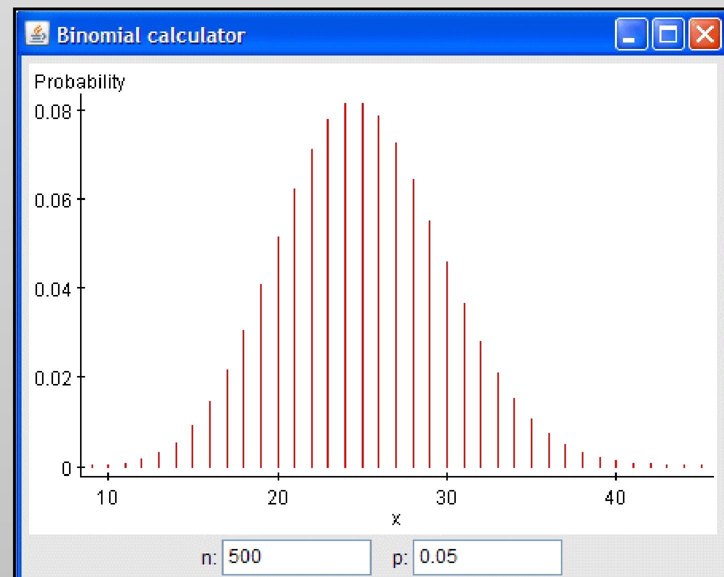
asymmetric when p≠0.5

but, as n increases, approaches symmetry even with very low or very high p









in the examples on page 29 and page 32 …

- *fixed number of trials*

  5 births on page 29, 10 births on page 32

- *trials are independent*

  probability of any particular birth being low birth weight infant has nothing to do with the probability of any other birth being low birth weight

- *each trial has only two possible outcomes*

  low birth weight, not low birth weight

- *probabilities remain constant for each trials*

  probability of low birth weight is the same for each birth

EXPECTED VALUE AND VARIANCE OF THE BINOMIAL DISTRIBUTION

		from Rosner …	example on page 18 14 births, P(girl) = 0.50 …
expected value	μ	np	$14 \times 0.5 = 7$

(logical given the frequency definition of … *p = k / n*)

variance	σ^2	npq	$14 \times 0.5 \times 0.5 = 3.5$
standard deviation	σ	$\sqrt{npq}$	$\sqrt{14 \times 0.5 \times 0.5} = 1.9$

---

same statistics calculated earlier (on page 19) as…

mean	$E(X) = \mu = \sum x \Pr(X=x)$	$\mu = 6.993 \approx 7.0$
variance	$Var(X) = \sigma^2 = \sum (x - \mu)^2 \Pr(X=x)$	$\sigma^2 = 3.564 \approx 3.6$
standard deviation		$\sigma = 1.9$

CALCULATING BINOMIAL PROBABILITIES

- FORMULA
- BINOMIAL TABLES
- STATCRUNCH
- STATDISK
- EXCEL
- SAS
- VARIOUS PAGES ON THE WEB

example 4.30 from Rosner …

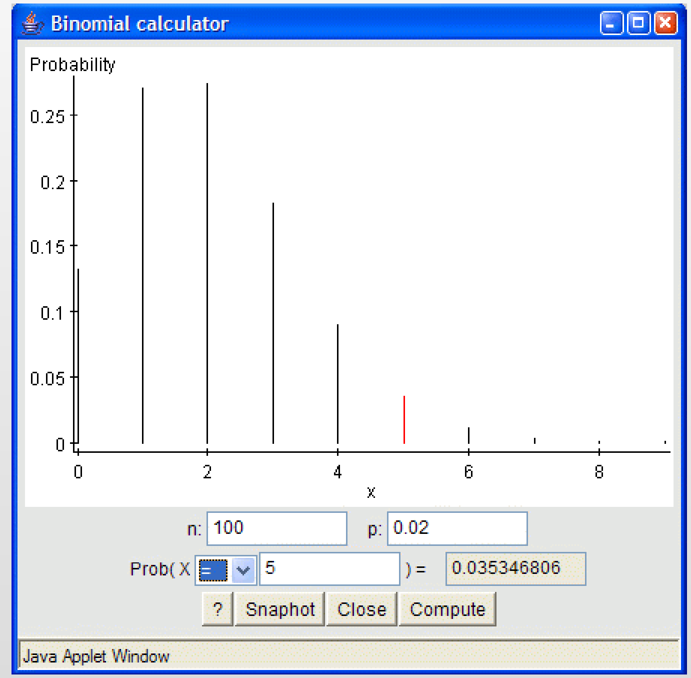
100 men aged 60-64 receive flu vaccine in 1986 and 5 of them die … is this an unusual event

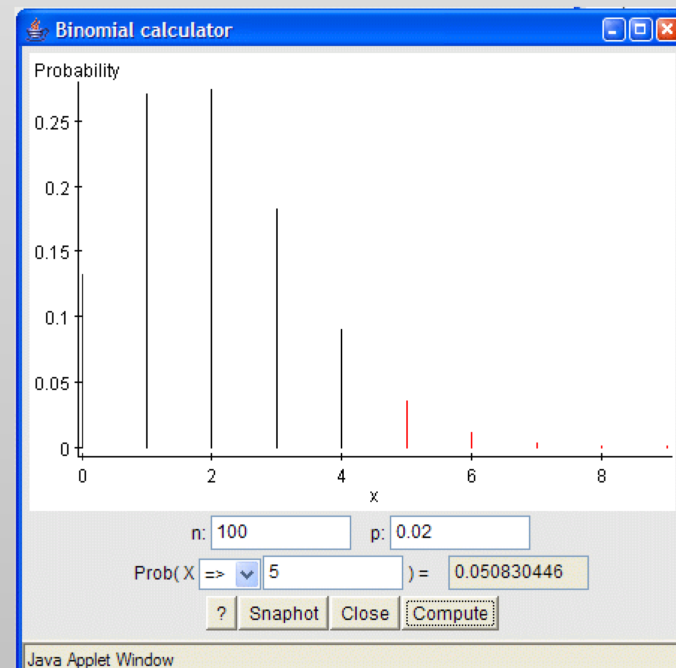
need P(death) for 60-64 year old men … from 1986 US data, P(death) = 0.02

use StatCrunch … results on right …

P(X=5) = 0.035
P(X≥5) = 0.051 (same answer as Rosner)

use X≥5 and conclude … it's close !





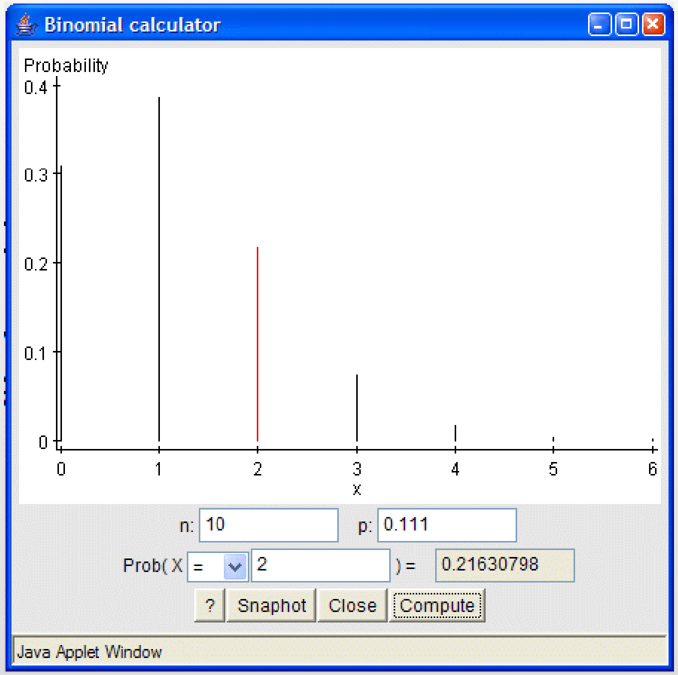
review question 4C1 from Rosner ...

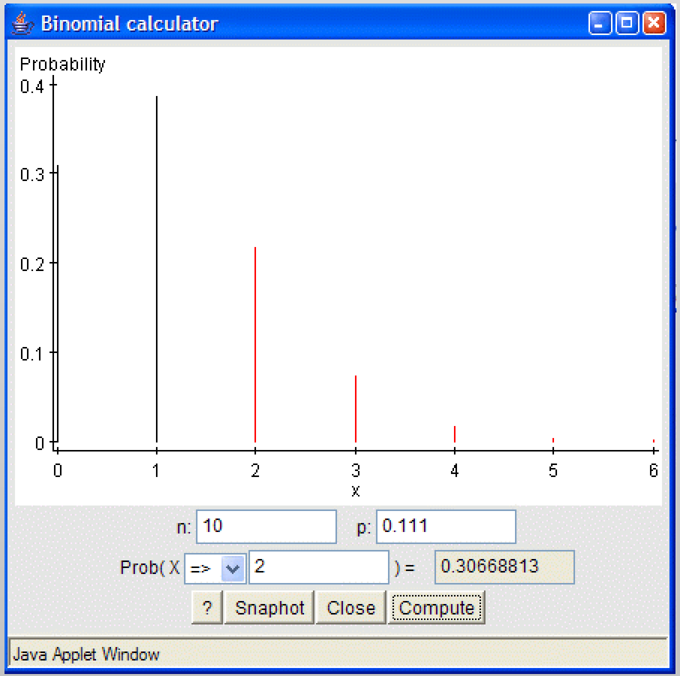
lifetime probability of a woman developing breast cancer ... P(breast cancer) = 1/9 = 0.111

what is the probability that among 10 women, exactly 2 develop breast cancer ... at least 2 develop breast cancer

use StatCrunch ... results on right ...

$P(X=2) = 0.216$
$P(X\geq 2) = 0.307$





last two example done with Statdisk ... no graphics, but you get P(X=x), P(X≤x), and P(X≥x) all in the same table ...

example 4.3 from Rosner

Binomial Probability

Num Trials, n: 100

Success Prob, p: 0.02

Evaluate

Mean: 2.0000
St Dev: 1.4000
Variance: 1.9600

x	P(x)	P(x or fewer)	P(x or greater)
0	0.1326196	0.1326196	1.0000000
1	0.2706522	0.4032717	0.8673804
2	0.2734139	0.6766856	0.5967283
3	0.1822759	0.8589616	0.3233144
4	0.0902080	0.9491696	0.1410384
5	0.0353468	0.9845164	0.0508304
6	0.0114216	0.9959379	0.0154836
7	0.0031301	0.9990681	0.0040621
8	0.0007426	0.9998107	0.0009319
9	0.0001549	0.9999656	0.0001893
10	0.0000288	0.9999944	0.0000344
11	0.0000048	0.9999992	0.0000056
12	0.0000007	0.9999999	0.0000008
13	0.0000001	1.0000000	0.0000001
14	0.0000000	1.0000000	0.0000000

Help ? Clear Copy

review question 4C1 from Rosner

Binomial Probability

Num Trials, n: 10

Success Prob, p: 0.111

Evaluate

Mean: 1.1100
St Dev: 0.9934
Variance: 0.9868

x	P(x)	P(x or fewer)	P(x or greater)
0	0.3083313	0.3083313	1.0000000
1	0.3849806	0.6933119	0.6916687
2	0.2163080	0.9096199	0.3066881
3	0.0720216	0.9816414	0.0903801
4	0.0157370	0.9973784	0.0183586
5	0.0023579	0.9997363	0.0026216
6	0.0002453	0.9999817	0.0002637
7	0.0000175	0.9999992	0.0000183
8	0.0000008	1.0000000	0.0000008
9	0.0000000	1.0000000	0.0000000
10	0.0000000	1.0000000	0.0000000

Help ? Clear Copy

example 4C1 from Rosner using a binomial table ...

the table on the right is a portion of a binomial table, similar to Table 1 in the Rosner appendix ...it is easy to find ... n=10, k=2 ... however, there is no p = 0.1111

if you wanted to use this table, you would have to interpolate ... with n=10, k=2 ...

at p=0.1, P(X=2) = 0.194
at p=0.2, P(X=2) = 0.302

if you interpolate assuming p=0.11 (not 0.1111), you would say that P(X=2) = 0.205, close to the exact answer from Statcrunch or Statdisk

2 BINOMIAL PROBABILITIES

Binomial probabilities:
$\binom{n}{x} p^x (1-p)^{n-x}$

n	x	0.1	0.2	0.25
8	0	0.430	0.168	0.100
	1	0.383	0.336	0.267
	2	0.149	0.294	0.311
	3	0.033	0.147	0.208
	4	0.005	0.046	0.087
	5	0.000	0.009	0.023
	6	0.000	0.001	0.004
	7	0.000	0.000	0.000
	8	0.000	0.000	0.000
9	0	0.387	0.134	0.075
	1	0.387	0.302	0.225
	2	0.172	0.302	0.300
	3	0.045	0.176	0.234
	4	0.007	0.066	0.117
	5	0.001	0.017	0.039
	6	0.000	0.003	0.009
	7	0.000	0.000	0.001
	8	0.000	0.000	0.000
	9	0.000	0.000	0.000
10	0	0.349	0.107	0.056
	1	0.387	0.268	0.188
	2	0.194	0.302	0.282
	3	0.057	0.201	0.250
	4	0.011	0.088	0.146
	5	0.001	0.026	0.058
	6	0.000	0.006	0.016
	7	0.000	0.001	0.003
	8	0.000	0.000	0.000
	9	0.000	0.000	0.000
	10	0.000	0.000	0.000

... if you forgot how or never learned how to interpolate, don't bother ... in real life, you will use software, not a table ...

POISSON PROBABILITY DISTRIBUTION

$$P(X = x) = \frac{\mu^x e^{-\mu}}{x!}$$ where $e \approx 2.71828$

random variable X is the number of occurrences of an event over some interval ... μ is the expected number of events over that interval ...
x is the observed number of events over that interval

... the Poisson probability formula is applicable to situations that meet the following criteria ...

- occurrences are random
- occurrences are independent
- occurrences are uniformly distributed over the interval

CHARACTERISTICS

- associated with rare events
- mean and variance are the same … μ is estimated as the number of events that occur over a fixed interval … the interval can be time, distance, area, volume, etc.
- not like the binomial in that there is no consideration of sample size or probability of success --- depends only on μ, a count
- approximates the binomial with a large number of trials and a small probability of success
- NOTE: binomial is a fixed number of trials … Poisson is a fixed interval

EXPECTED VALUE AND VARIANCE OF THE POISSON DISTRIBUTION

from Rosner … equation 4.9 … for a Poisson distribution with parameter μ, the mean and variance are both equal to μ

estimate μ with the observed number of events over a fixed time interval

Table 4.9 in Rosner

AN EXAMPLE OF USING THE FORMULA

example 4.2 from Rosner ... 12 cases of leukemia are found in town where 6 are expected ... does that town have an excess number of leukemia cases

assume leukemia is a rare event ... all you have are counts ... there in no fixed number of trials ... there is no probability of leukemia given

though not stated, these counts are likely over a fixed time interval ... that interval is 10-years (from problems 4.54-4.57 in the chapter 4 study guide)

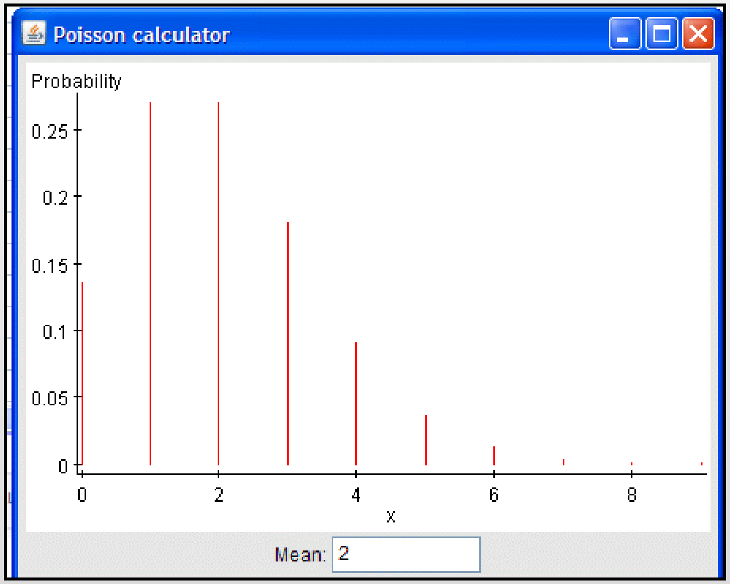
$$P(X = 12) = \frac{6^{12} e^{-6}}{12!} = 0.01126$$

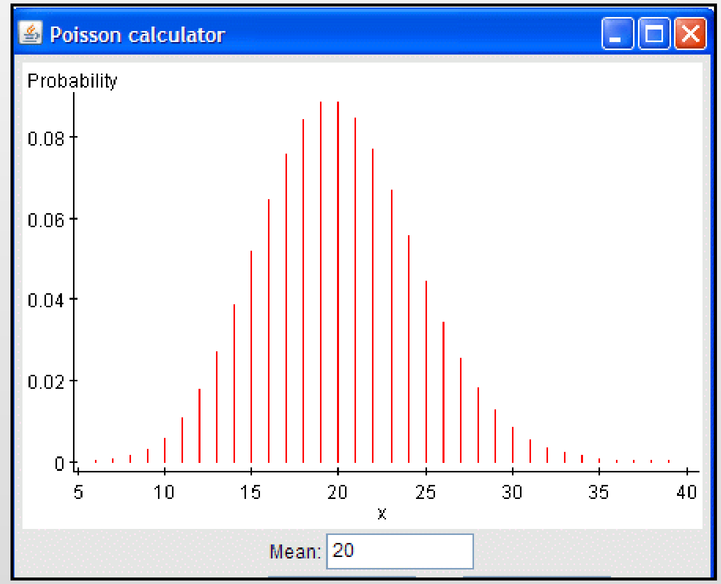
you would also have to compute P(X=13), P(X=14), etc. and add them all to compute $P(X \geq 12)$

you will NEVER use that formula

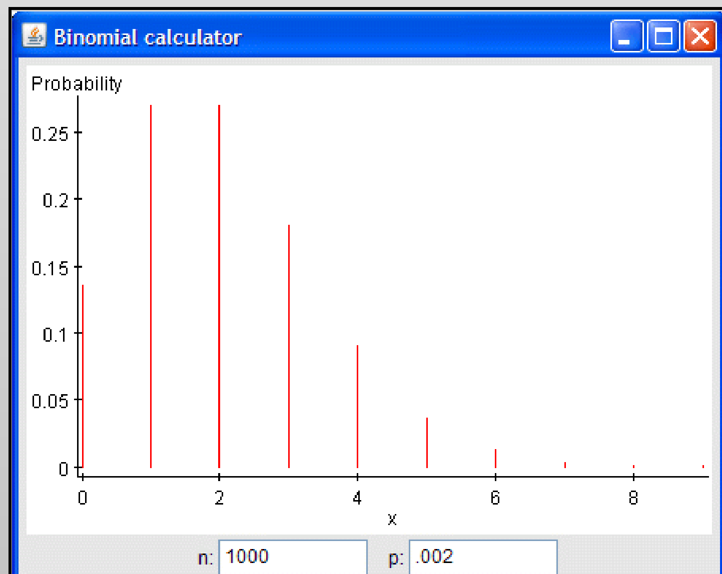
SOME POISSON DISTRIBUTIONS

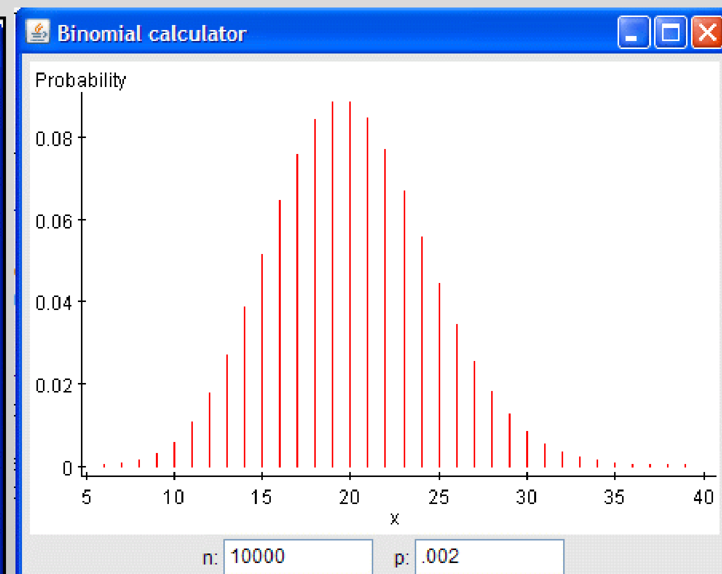
asymmetric with low μ … as μ increases, symmetry increases





binomial distributions with the same mean (np) as the Poisson distributions above them





CALCULATING POISSON PROBABILITIES

- FORMULA
- POISSON TABLES
- STATCRUNCH
- STATDISK
- EXCEL
- SAS
- VARIOUS PAGES ON THE WEB

again ... example 4.2 from Rosner ...

12 cases of leukemia are found in town where 6 are expected … does that town have an excess number of leukemia cases

Poisson table on the right ... similar to Table 2 in the Rosner appendix … mean (μ) across the top

	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6
0	0.00610	0.00552	0.00499	0.00452	0.00409	0.00370	0.00335	0.00303	0.00274	0.00248
1	0.03109	0.02869	0.02646	0.02439	0.02248	0.02071	0.01907	0.01756	0.01616	0.01487
2	0.07929	0.07458	0.07011	0.06585	0.06181	0.05798	0.05436	0.05092	0.04768	0.04462
3	0.13479	0.12928	0.12386	0.11853	0.11332	0.10823	0.10327	0.09845	0.09377	0.08924
4	0.17186	0.16806	0.16411	0.16002	0.15582	0.15153	0.14717	0.14276	0.13831	0.13385
5	0.17529	0.17479	0.17396	0.17282	0.17140	0.16971	0.16777	0.16560	0.16321	0.16062
6	0.14900	0.15148	0.15366	0.15554	0.15712	0.15840	0.15938	0.16008	0.16049	0.16062
7	0.10856	0.11253	0.11634	0.11999	0.12345	0.12672	0.12978	0.13263	0.13527	0.13768
8	0.06921	0.07314	0.07708	0.08099	0.08487	0.08870	0.09247	0.09616	0.09976	0.10326
9	0.03922	0.04226	0.04539	0.04859	0.05187	0.05519	0.05856	0.06197	0.06540	0.06884
10	0.02000	0.02198	0.02406	0.02624	0.02853	0.03091	0.03338	0.03594	0.03859	0.04130
11	0.00927	0.01039	0.01159	0.01288	0.01426	0.01573	0.01730	0.01895	0.02070	0.02253
12	0.00394	0.00450	0.00512	0.00580	0.00654	0.00734	0.00822	0.00916	0.01018	0.01126
13	0.00155	0.00180	0.00209	0.00241	0.00277	0.00316	0.00360	0.00409	0.00462	0.00520
14	0.00056	0.00067	0.00079	0.00093	0.00109	0.00127	0.00147	0.00169	0.00195	0.00223
15	0.00019	0.00023	0.00028	0.00033	0.00040	0.00047	0.00056	0.00065	0.00077	0.00089

look in the last column ...

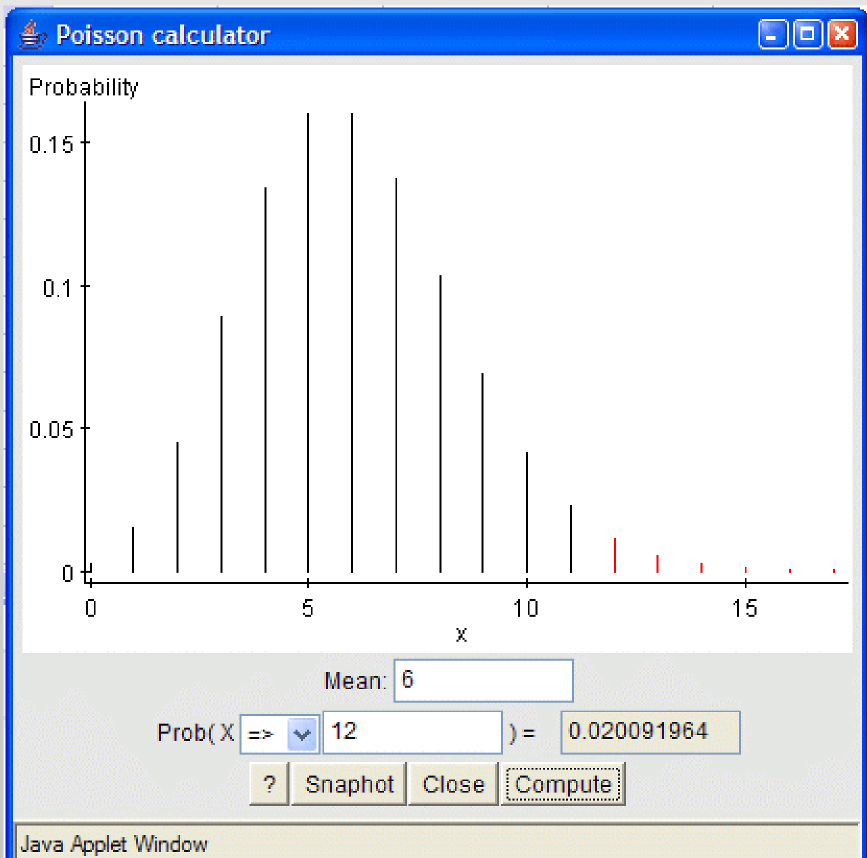
$P(X=12) = 0.01126$
$P(X=13) = 0.00520$
$P(X=14) = 0.00223$
$P(X=15) = 0.00089$

$P(X \geq 12)$ = sum of the above = 0.01958

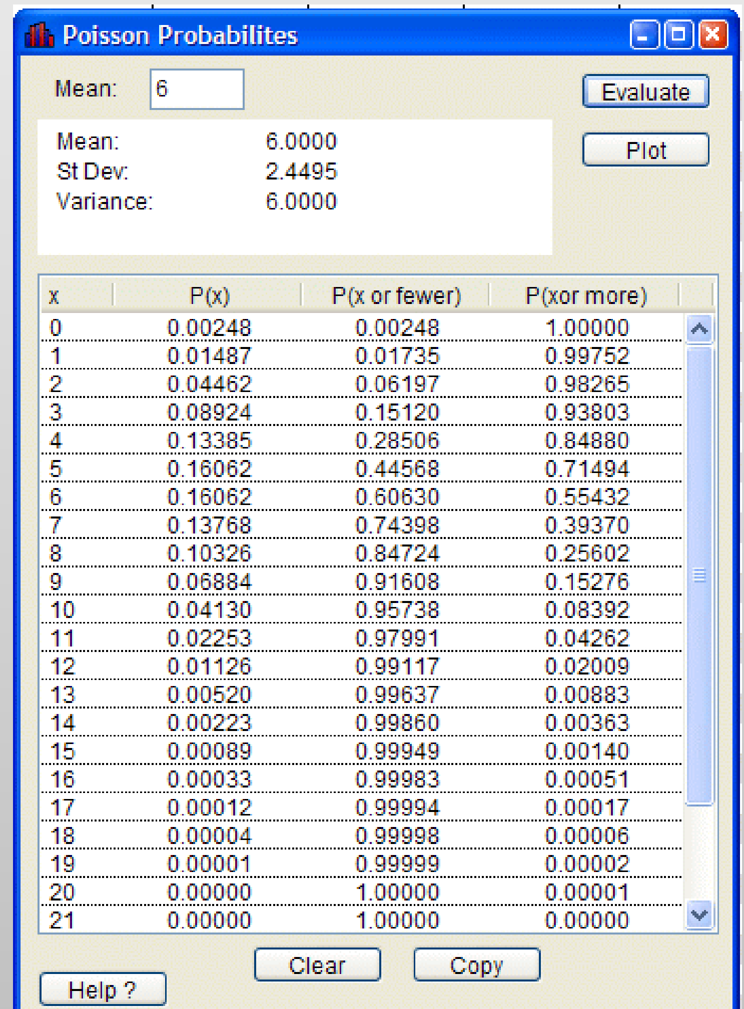
$P(X \geq 12) < 0.05$ what do you conclude use StatCrunch and Statdisk (column 3) ... $P(X \geq 12) = 0.02009$ in both

differs from previous calculation of $P(X \geq 12)=0.01958$ using the table since we stopped at $P(X=15)$ and you can see in Statdisk (column 1) that we should have continued to $P(X=19)$

StatCrunch ...



Statdisk...



Poisson Probabilites

Mean: 6

Mean:	6.0000
St Dev:	2.4495
Variance:	6.0000

x	P(x)	P(x or fewer)	P(xor more)
0	0.00248	0.00248	1.00000
1	0.01487	0.01735	0.99752
2	0.04462	0.06197	0.98265
3	0.08924	0.15120	0.93803
4	0.13385	0.28506	0.84880
5	0.16062	0.44568	0.71494
6	0.16062	0.60630	0.55432
7	0.13768	0.74398	0.39370
8	0.10326	0.84724	0.25602
9	0.06884	0.91608	0.15276
10	0.04130	0.95738	0.08392
11	0.02253	0.97991	0.04262
12	0.01126	0.99117	0.02009
13	0.00520	0.99637	0.00883
14	0.00223	0.99860	0.00363
15	0.00089	0.99949	0.00140
16	0.00033	0.99983	0.00051
17	0.00012	0.99994	0.00017
18	0.00004	0.99998	0.00006
19	0.00001	0.99999	0.00002
20	0.00000	1.00000	0.00001
21	0.00000	1.00000	0.00000

Evaluate · Plot · Clear · Copy · Help ?

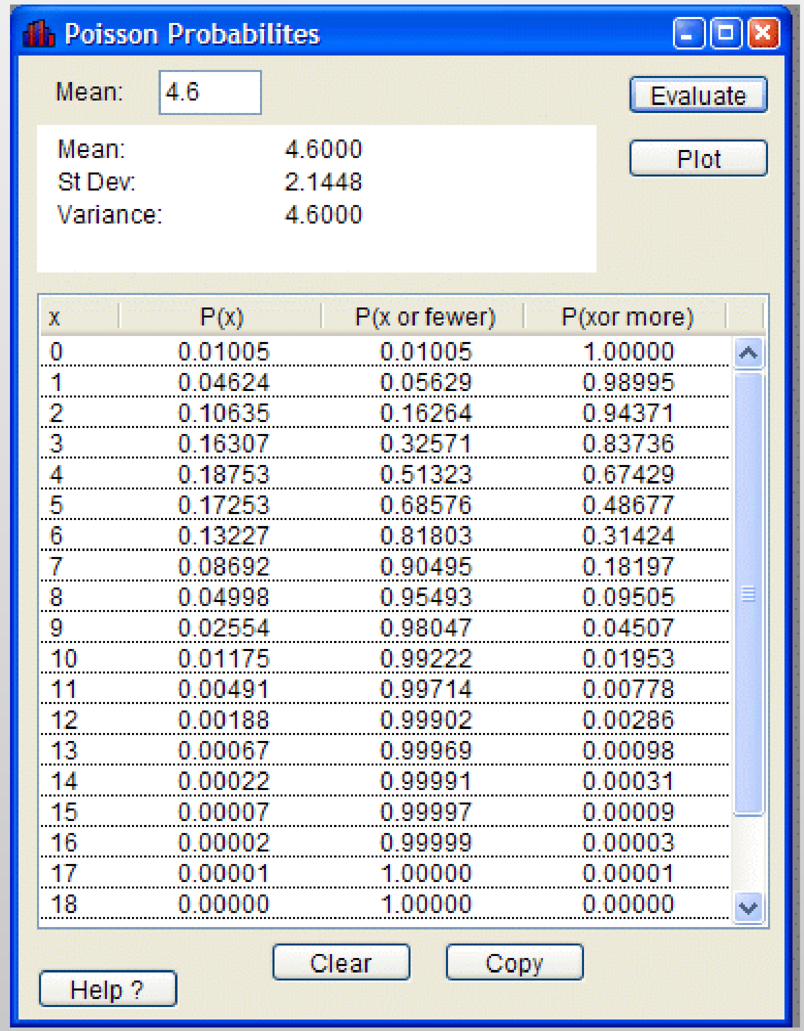
example 4.36 from Rosner …

if the expected number of deaths from typhoid fever in a one year period is 4.6, calculate the probability distribution … using Statdisk … table on right

how many deaths would be a rare event (p<0.05)

since the probability of 9 or more deaths is < 0.05, 9 or more deaths is rare event

0 deaths is also a rare event



x	P(x)	P(x or fewer)	P(xor more)
0	0.01005	0.01005	1.00000
1	0.04624	0.05629	0.98995
2	0.10635	0.16264	0.94371
3	0.16307	0.32571	0.83736
4	0.18753	0.51323	0.67429
5	0.17253	0.68576	0.48677
6	0.13227	0.81803	0.31424
7	0.08692	0.90495	0.18197
8	0.04998	0.95493	0.09505
9	0.02554	0.98047	0.04507
10	0.01175	0.99222	0.01953
11	0.00491	0.99714	0.00778
12	0.00188	0.99902	0.00286
13	0.00067	0.99969	0.00098
14	0.00022	0.99991	0.00031
15	0.00007	0.99997	0.00009
16	0.00002	0.99999	0.00003
17	0.00001	1.00000	0.00001
18	0.00000	1.00000	0.00000

review question 4D1 from Rosner ...

number of motor vehicle deaths in a city is Poisson distributed with an average of 8 fatalities per week

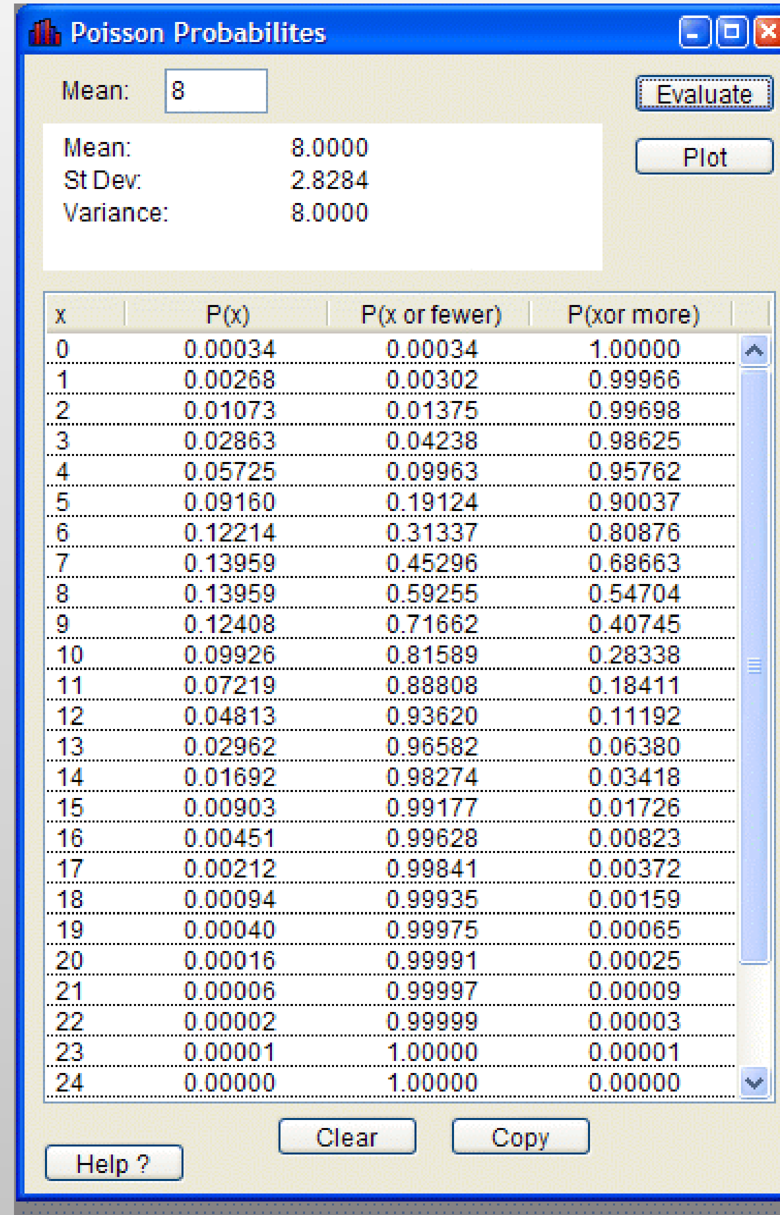
what is the probability of 12 deaths in week ... of 12 or more deaths in a week

how many deaths constitute an unusually high number of deaths in a week

using Statdisk ... table on right

exactly 12 $P(X=12) = 0.04813$
12 or more $P(X\geq 12) = 0.11192$

what's unusual $P(X\geq 13) = 0.06380$
$P(X\geq 14) = 0.03418$
must be 14 or more



x	P(x)	P(x or fewer)	P(xor more)
0	0.00034	0.00034	1.00000
1	0.00268	0.00302	0.99966
2	0.01073	0.01375	0.99698
3	0.02863	0.04238	0.98625
4	0.05725	0.09963	0.95762
5	0.09160	0.19124	0.90037
6	0.12214	0.31337	0.80876
7	0.13959	0.45296	0.68663
8	0.13959	0.59255	0.54704
9	0.12408	0.71662	0.40745
10	0.09926	0.81589	0.28338
11	0.07219	0.88808	0.18411
12	0.04813	0.93620	0.11192
13	0.02962	0.96582	0.06380
14	0.01692	0.98274	0.03418
15	0.00903	0.99177	0.01726
16	0.00451	0.99628	0.00823
17	0.00212	0.99841	0.00372
18	0.00094	0.99935	0.00159
19	0.00040	0.99975	0.00065
20	0.00016	0.99991	0.00025
21	0.00006	0.99997	0.00009
22	0.00002	0.99999	0.00003
23	0.00001	1.00000	0.00001
24	0.00000	1.00000	0.00000

the leukemia problem in different words (from the study guide) ... A study was performed in Woburn, MA, looking at the rate of leukemia among children ($\leq$ age 19) in the community in comparison to statewide leukemia rates. Suppose there arc 12,000 children in the community that have lived there for a 10-year period and 12 leukemias have occurred in 10 years.

4.54 If the statewide incidence rate of leukemia in children is 5 events per 100,000 children per year (i.e., per 100,000 person-years) then how many leukemias would be expected in Woburn over the 10-year period if the statewide rates were applicable?

how would you calculate μ given the above information

4.55 What is the probability of obtaining exactly 12 events over a 10-year period if statewide incidence rates were applicable?

4.56 What is the probability of obtaining at least 12 events over the 10-year period if the statewide incidence rates were applicable?

how would you use Table 2 in the Rosner appendix to answer the above two questions

4.57 How do you interpret the results in Problem 4.56?

POISSON APPROXIMATION TO THE BINOMIAL

- the Poisson distribution approximates the binomial with large n and small P
- conservative rules ... large n is 100+ and small P is $\leq$0.01
- rationale … Poisson easier to work with than the binomial
- DUH … we have computers and software
- NO REASON TO USE AN APPROXIMATION IF YOU HAVE N AND P AND THE CRITERIA FOR USING THE BINOMIAL (ON PAGE 31) ARE MET

***** EXTRA MATERIAL *****

identifying unusual results with probability (from Triola)

do not use P(X=x) to identify 'significant' probabilities (P<0.05)

use either P(X≤x) or P(X≥x) ... cumulative probabilities

example ... coin tossing ... you are trying to determine if you have a 'fair' coin, one where...

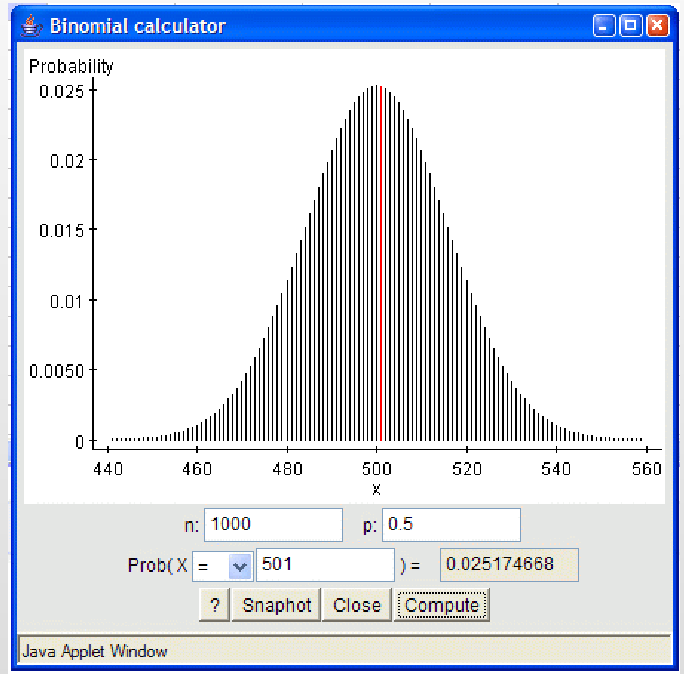
P(X=heads) = P(X=tails) = 0.50

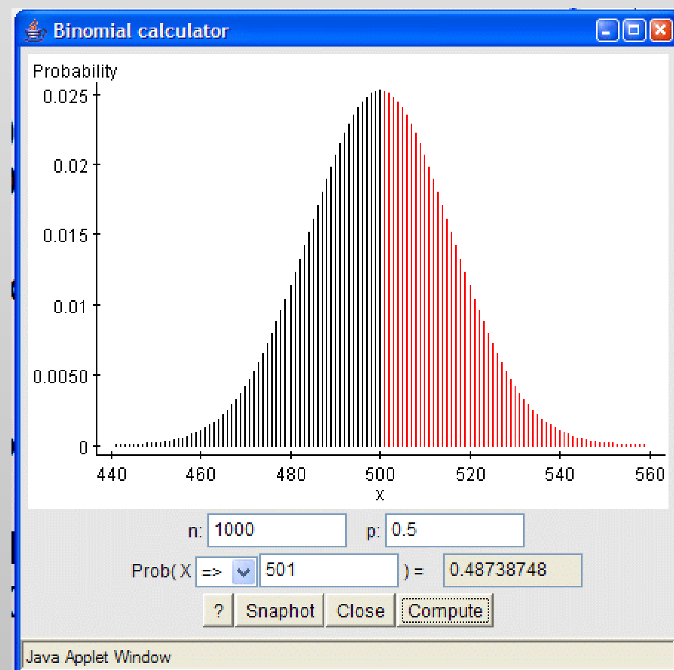
in 1,000 tosses, you get 501 heads ... is the coin fair

P(X=501) = 0.025 rare event
P(X≥501) = 0.487 not a rare event

which of the above probabilities would you use

which matches what you would conclude without calculating a probability





identifying unusual results with probability (from Rosner)

example ... random event is a household with both parents having bronchitis ... and you know that ...

P(X=x) = 0.05 (in 5% of households, both parents have bronchitis)

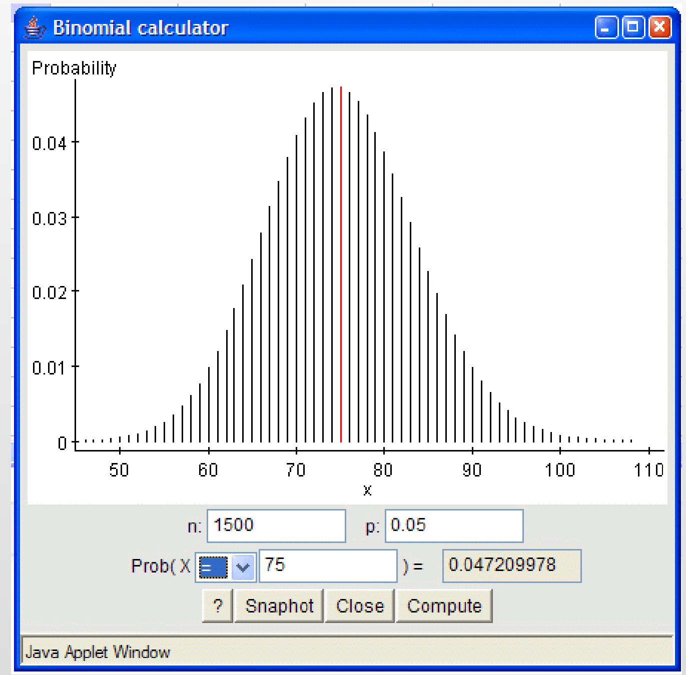
you survey 1,500 households and find 75 where both parents have bronchitis ... is this different from what was expected

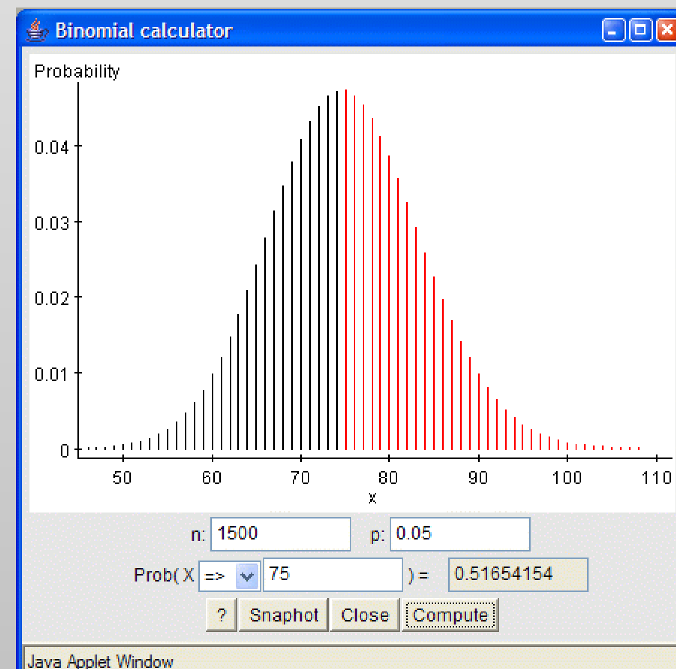
P(X=75) = 0.047 rare event
P(X≥75) = 0.517 not a rare event

in your sample 5% (75 / 1500) of households, matching what you already know about the percentage of households where two parents have bronchitis

which of the above probabilities would you use

which matches what you would conclude without calculating a probability





A PERMUTATION EXAMPLE

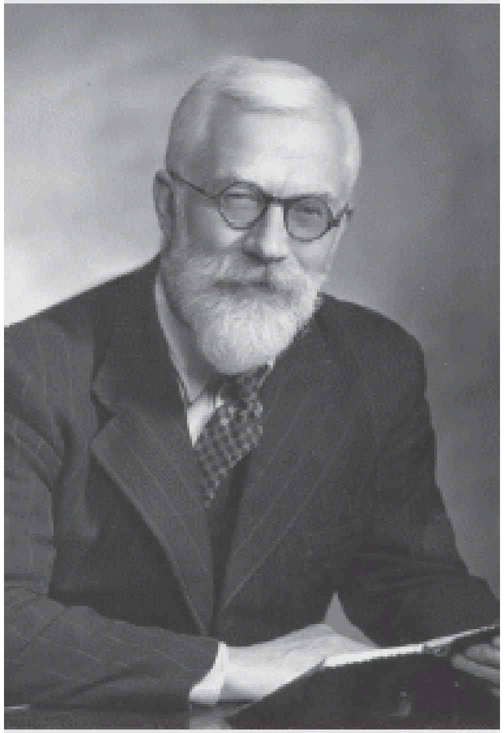
there are 3 patients with a rare disease and 2 of them are to be chosen to participate in an experiment ... the first patient chosen will get drug A, the second one chosen gets drug B

since ORDER DOES MATTER in this selection (order determines what drug you get), this is a PERMUTATION

how many different ways can the 2 people be chosen ...

1 and 2	1 and 3
2 and 1	2 and 3
3 and 1	3 and 2

or ... $${}_nP_k = \frac{n!}{(n-k)!} = \frac{3!}{(3-2)!} = \frac{3 \times 2 \times 1}{1} = 6$$



The great R. A. Fisher wrote in 1926: "Personally, the writer prefers to set a low standard of significance at the 5 percent point, and ignore entirely all results which fail to reach that level. A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance." (Quoted in Moore, 1979 edition).

It is the fate of a guru that what he sees as a convenient but arbitrary option is taken by followers as written in stone. But it is a philosophy that must be abandoned.

where does the magic of $P<0.05$ come from

it started with R. A. Fisher (on the left) ...

"...the writer prefers to set a low standard of significance at the 5 percent point..."

and ... note the similarity to Los Angeles Lakers coach Phil Jackson ... is that a "rare event"